

IN THE CLAIMS

1. (Previously Presented) A method comprising:

representing an input document image with a sequence of template identifiers;

replacing the template identifiers with alphabet characters according to language statistics to generate a text string representing the input document image;

searching, among a plurality of documents in a database, for at least one of the plurality of documents that matches the input document based on the text string; and

examining whether the at least one matched document satisfies a predetermined security criteria based on an attribute associated with the at least one matched document, to determine whether an operation on the input document is allowed.

2. – 40. (Cancelled)

41. (Original) The method of claim 1, wherein determining whether the at least one matched document satisfies a predetermined security criteria comprises determining whether the at least one matched document is a confidential document that requires an authorization before operating on the input document.

42. (Original) The method of claim 41, wherein if the at least one matched document is determined to be a confidential document, the method further comprises:

prompting a user for an authorization; and

determining whether the authorization received from the user satisfies a level of the authorization required by the at least one matched document.

43. (Original) The method of claim 1, wherein the plurality of documents comprises a hierarchy of confidential documents, at least a portion of the confidential documents being associated with a respective confidentiality rating that requires a specific authorization associated with the respective rating.

44. (Original) The method of claim 1, wherein the operation on the input document comprises at least one of scanning and copying the input document.

45. (Original) The method of claim 1, wherein the operation on the input document comprises printing the at least one matched document.

46. (Original) The method of claim 1, wherein determining whether the at least one matched document satisfies a predetermined security criteria comprises determining whether the at least one matched document is a copyright protected document before copying the input document.

47. (Original) The method of claim 46, wherein if the at least one matched document is a copyright protected document, the method further comprises identifying a copyright holder and a copyright license fee associated with the copyright protected document.

48. (Original) The method of claim 47, further comprising:
recording an identity of a user submitting the input document, dates, and number of copy operations performed on the input document; and

storing the identify of the user, dates, and the number of copy operations in the database for accounting purposes.

49. (Original) The method of claim 48, further comprising transmitting the identity of the user and the number of copies of the input document to a remote facility over a network for billing purposes.

50. (Original) The method of claim 1, wherein the input document is a symbolically compressed document and the plurality of documents including the at least one matched document are stored as symbolically compressed documents.

51. (Original) The method of claim 1, further comprising mapping the alphabet characters to the template identifiers based at least partly on frequency of occurrence of the template identifiers.

52. (Original) The method of claim 1, further comprising extracting n-gram indexing terms from the text string, wherein the comparison of the input document and the plurality of documents is performed based on the n-gram indexing terms.

53. (Original) The method of claim 52, wherein extracting n-gram indexing terms comprises:

selecting alphabet characters from the text string that satisfy a predicate; and
combining the selected alphabet characters to form n-grams, n being an integer.

54. (Original) A document processing system comprising:

a deciphering module to generate a first text string on a sequence of template identifiers in a first document and to generate a second text string based on a sequence of template identifiers in a second document;

a comparison module to generate a measure of similarity between the first and the second documents based on the first and second text strings to determine whether the first and second documents are matched; and

a security module to examine whether the second document satisfies a predetermined security criteria based on an attribute associated with the second document to determine whether an operation on the first document is allowed.

55. (Original) The document processing system of claim 54, wherein the security module further determines whether the second document is a confidential document that requires an authorization before operating on the first document.

56. (Original) The document processing system of claim 55, wherein if the second document is determined to be a confidential document, the security module further prompts a user for an authorization, and determines whether the authorization received from the user satisfies a level of the authorization required by the second document.

57. (Original) The document processing system of claim 54, wherein the second document is a member of a hierarchy of confidential documents stored in a database, at least a portion of the confidential documents being associated with a respective

confidentiality rating that requires a specific authorization associated with the respective rating.

58. (Original) The document processing system of claim 54, wherein the operation on the first document comprises at least one of scanning and copying the first document.

59. (Original) The document processing system of claim 54, wherein the operation on the first document comprises printing the second document.

60. (Original) The document processing system of claim 54, wherein the security module further determines whether the second document is a copyright protected document before copying the first document.

61. (Original) The document processing system of claim 60, wherein if the second document is a copyright protected document, the security module further identifies a copyright holder and a copyright license fee associated with the copyright protected document.

62. (Original) The document processing system of claim 61, further comprising an accounting module to:

record an identity of a user submitting the first document, dates, and number of copy operations performed on the first document, and

store the identify of the user, dates, and the number of copy operations in a database for accounting purposes.

63. (Original) The document processing system of claim 62, further comprising a communication module to transmit the identity of the user and the number of copies of the first document to a remote facility over a network for billing purposes.

64. (Original) The document processing system of claim 54, wherein the first and second documents are symbolically compressed documents.

65. (Original) The document processing system of claim 54, further comprising a conditional n-gram module coupled to receive the first and second text strings from the deciphering module, and to extract n-gram indexing terms from the text string, wherein the comparison of the first document and the second document is performed based on the n-gram indexing terms.

66. (Original) The document processing system of claim 65, wherein the n-gram indexing terms are extracted by:

selecting alphabet characters from the text string that satisfy a predicate; and
combining the selected alphabet characters to form n-grams, n being an integer.

67. (Original) The document processing system of claim 66, wherein the deciphering module maps the alphabet characters to the template identifiers based at least partly on frequency of occurrence of the template identifiers.

68. (Original) An article of manufacture including one or more computer-readable media that embody a program of instructions, when executed by one or more processors in the processing system, causes the one or more processors to performing a method, the method comprising:

generating a text string from an input document image represented by a sequence of template identifiers;

replacing the template identifiers with alphabet characters according to language statistics to generate a text string representing the input document image;

searching, among a plurality of documents in a database, for at least one of the plurality of documents that matches the input document based on the text string; and

examining whether the at least one matched document satisfies a predetermined security criteria based on an attribute associated with the at least one matched document, to determine whether an operation on the input document is allowed.

69. (Original) The article of claim 68, wherein determining whether the at least one matched document satisfies a predetermined security criteria comprises determining whether the at least one matched document is a confidential document that requires an authorization before operating on the input document.

70. (Original) The article of claim 69, wherein if the at least one matched document is determined to be a confidential document, the method further comprises:

prompting a user for an authorization; and

determining whether the authorization received from the user satisfies a level of the authorization required by the at least one matched document.

71. (Original) The article of claim 68, wherein the plurality of documents comprises a hierarchy of confidential documents, at least a portion of the confidential documents being associated with a respective confidentiality rating that requires a specific authorization associated with the respective rating.

72. (Original) The article of claim 68, wherein the operation on the input document comprises at least one of scanning and copying the input document.

73. (Original) The article of claim 68, wherein the operation on the input document comprises printing the at least one matched document.

74. (Original) The article of claim 68, wherein determining whether the at least one matched document satisfies a predetermined security criteria comprises determining whether the at least one matched document is a copyright protected document before copying the input document.

75. (Original) The article of claim 74, wherein if the at least one matched document is a copyright protected document, the method further comprises identifying a copyright holder and a copyright license fee associated with the copyright protected document.

76. (Original) The article of claim 75, wherein the method further comprises: recording an identity of a user submitting the input document, dates, and number of copy operations performed on the input document; and

storing the identify of the user, dates, and the number of copy operations in the database for accounting purposes.

77. (Original) The article of claim 76, wherein the method further comprises transmitting the identity of the user and the number of copies of the input document to a remote facility over a network for billing purposes.

78. (Original) The article of claim 54, wherein the input document is a symbolically compressed document and the plurality of documents including the at least one matched document are stored as symbolically compressed documents.

79. (Original) The article of claim 54, wherein the method further comprises mapping the alphabet characters to the template identifiers based at least partly on frequency of occurrence of the template identifiers.

80. (Original) The article of claim 54, wherein the method further comprises extracting n-gram indexing terms from the text string, wherein the comparison of the input document and the plurality of documents is performed based on the n-gram indexing terms.

81. (Original) The article of claim 80, wherein extracting n-gram indexing terms comprises:

selecting alphabet characters from the text string that satisfy a predicate; and
combining the selected alphabet characters to form n-grams, n being an integer.